

Estimation of crop water requirement based on principal component analysis and geographically weighted regression

WANG JingLei^{1,2}, KANG ShaoZhong^{1,3*}, SUN JingSheng² & CHEN ZhiFang²

¹ College of Water Resources and Architectural Engineering, Northwest Agricultural & Forestry University, Yangling 712100, China;

² Farmland Irrigation Research Institute, Chinese Academy of Agricultural Sciences and Ministry of Water Resources, Xinxiang 453003, China;

³ College Water Conservancy and Civil Engineering, China Agricultural University, Beijing 100083, China

Received August 31, 2012; accepted January 22, 2013; published online April 22, 2013

In this study the principal component analysis (PCA) and geographically weighted regression (GWR) are combined to estimate the spatial distribution of water requirement of the winter wheat in North China while the effect of the macro- and micro-topographic as well as the meteorological factors on the crop water requirement is taking into account. The spatial distribution characteristic of the water requirement of the winter wheat in North China and its formation are analyzed based on the spatial variation of the main affecting factors and the regression coefficients. The findings reveal that the collinearity can be effectively removed when PCA is applied to process all of the affecting factors. The regression coefficients of GWR displayed a strong variability in space, which can better explain the spatial differences of the effect of the affecting factors on the crop water requirement. The evaluation index of the proposed method in this study is more efficient than the widely used Kriging method. Besides, it could clearly show the effect of those affecting factors in different spatial locations on the crop water requirement and provide more detailed information on the region where those factors suddenly change. To sum up, it is of great reference significance for the estimation of the regional crop water requirement.

principal component analysis, geographically weighted regression, crop water requirement, spatial variability, kriging

Citation: Wang J L, Kang S Z, Sun J S, et al. Estimation of crop water requirement based on principal component analysis and geographically weighted regression. *Chin Sci Bull*, 2013, 58: 3371–3379, doi: 10.1007/s11434-013-5750-1

Crop water requirement is an important parameter in the calculation of the equilibrium of water and soil resources and the design, operation, and management of the irrigation projects [1]. Many methods of measuring and estimating crop water requirement have been proposed in China and foreign countries [2–5]. Although these methods exhibit accuracy in their respective applications to some extent, what is obtained when they are used are mostly point data, which can neither be used directly for other points nor as the average data of a large area [6]. Therefore, how to make full use of the data obtained from the measured points has been the focus of most studies since it is impossible to lay out as many test sites as desired [7–12]. Spatial interpolation methods now commonly used, such as inverse distance

method, are simple and easy to apply. However, only the effect of the geographic coordinates are considered in those methods and the change in the spatial structure of the crop water requirement is scarcely revealed. Although adjusted geological statistics is usually applied in the estimation of crop water requirements, the sample data should meet important prerequisites [13]. When regression analysis is adopted, both the effect of macro-terrain and that of the micro-terrain factors on crop water requirement are taken into account. However, those factors are complicated and a correlation among those factors is also found [14]. The correlation contradicts the hypothesis that all of these factors are independent from traditional global regression analysis. Moreover, crop water requirements are typically regionalized variables [8] and almost spatially dependent. In most traditional statistical analyses, spatial dependence also con-

*Corresponding author (email: kangshaozhong@tom.com)

tradicts the basic hypothesis that the observed values are independent [15]. Therefore, the classical global regression model can neither show the effect of those factors on crop water requirements in different places nor indicate the characteristics of the changes in crop water requirement locally or in different spatial locations because the regression coefficient provided by this model is considered constant, although this model can be used to analyze and describe the change in crop water requirement.

Principal component analysis (PCA) is used to group the factors into few integrated factors after the internal structure of their correlation matrix is studied. The main information of the primary factors that affect crop water requirement can also be preserved by grouping these primary factors, and their correlation can be reduced [16]. This method is currently applied in the component analysis of groundwater pollution [17], spatial distribution of the Mongolian gerbil [18], and determination of the principal factors in crop water requirements [19]. Geographically weighted regression (GWR), proposed by Fotheringham et al. [20], can be used to estimate the local linear regression at each point by using the distance weight of the sub-sample data and detect the variability as well as the spatial dependence of the spatial data because each parameter estimation is based on the distance weight of the sub-samples in adjacent observations [21]. This method is widely applied to analyze the spatial distribution of the rainfall in a complicated region [22], assess regional disasters [23], determine the effect of the society and environment on urbanization and land use [24], and evaluate the efficiency of irrigation [25]. This method is more effective than the traditional global regression analysis. This study is designed to demonstrate the effect of the principal factors on crop water requirement, improve the accuracy of crop water requirement estimation by combining PCA and GWR, and apply dimension reduction of the detection capability of PCA and GWR on spatial data variability.

1 Data source and method

1.1 Research scope and data source

The demand for water is higher than the water supply in North China, a key production base of winter wheat. This study chooses the winter wheat planting area in Henan Province and Shandong Province, Beijing and Tianjin municipalities, the southern part of the Great Wall in Hebei Province, and Shanxi Province except the spring wheat area in the northern part [26]. According to the division of the plantation, these areas ($59.74 \times 10^4 \text{ km}^2$) are mostly located in the North and Huanghai winter wheat areas, as well as in Xinyang in Henan Province located in the lower middle portions of the Yangtze River [27].

Meteorological data used in this study include the maximum temperature, the minimum temperature, the average temperature, the wind speed, the relative humidity, the pre-

cipitation, the evaporation, the sunshine duration, and the daily temperature range, which have been obtained from the China Meteorological Administration since 2004 (when the weather station was established). Based on the characteristics of the growing season of winter wheat at each station, the meteorological data are interpolated and processed to obtain the perennial average of the meteorological elements from 1961 to 2004. The water requirement data of winter wheat are obtained from two research projects, namely, the "Contour Map of National Major Crop Water Requirement" and the "National Irrigation Experiment Data Compilation". The data from these research projects are enriched and modified with the related research findings from a five-year period (such as the study on the water requirement of major crops in the Yellow River Basin sponsored by the Diversion of Southern Water to North Program). Measured data are used if available from the sites and years. If a site does not have any measured data in given year, then the water requirement of winter wheat is calculated by using the crop coefficient method, in which the reference crop evapotranspiration in 10 days is calculated by using the Penman-Monteith method, which is recommended by FAO-56. For the corresponding crop coefficient, the average crop coefficient is obtained from the experimental data of some test sites. Data on macro-terrain factors are extracted from the National Irrigation Experiment Database. Shuttle radar topographic mission 3 data are used in digital elevation models (DEM) after rigorous geometric calibration, registration, and projection transformation are performed. The micro-terrain factors, including slope, aspect, and shading degree, are derived from the Arcgis10.0 spatial analysis module based on DEM data of the study areas. Spatial processing, which was conducted by selecting the most appropriate method for each meteorological element, is performed for the comparative analysis of the interpolation methods provided by the Arcgis10.0 space analysis module.

1.2 Method

The factors that affect crop water requirement are correlated, which does not meet the requirement that those factors be independent in regression analysis. PCA can reduce a number of factors into several integrated ones by examining the internal structure of the factor matrix. Some independent factors that affect the crop water requirement, which is a linear combination of primary factors, can also be found [16]. This method improves data processing and operational efficiency, retains most information of the primary factors, and weakens the correlation among them. In this study, the ArcGIS software, a spatial analysis tool provided by the Environmental Systems Research Institute, Inc., is used to conduct PCA. The principal factors that affect crop water requirement can be determined by analyzing whether the effect of each factor on crop water requirement is significant or not at 1% significance level. Eq. (1) is used in this

study for the standard processing of the original data:

$$x_{ik} = \frac{x_{ik}^* - \bar{x}_k^*}{\sqrt{\text{var}(x_k^*)}}, \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots, p, \quad (1)$$

where, x_{ik} is the affecting factor after the standard processing and x_{ik}^* is the original data of the k affecting factor at the i monitoring point; \bar{x}_k^* and $\sqrt{\text{var}(x_k^*)}$ are the average and the standard deviation of the original data of k affecting factor, respectively; n is the total number of monitoring points; and p is the number of affecting factors.

Traditional global regression models assume that parameters are constant in the entire research scope. However, the parameters that affect the crop water requirement, e.g., the wind speed, are spatially variable. Therefore, it is difficult to use global regression models to process and analyze the spatial heterogeneity data, such as the crop water requirement. The GWR model proposed by Fotheringham et al. [20] allows certain changes in parameters in space by expanding the general linear regression model. The parameters of this model after expansion are the function of location i . The basic form of the GWR model is expressed as follows:

$$Y_i = \alpha_{i0} + \sum_{k=1}^p \alpha_{ik} x_{ik} + \theta_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where, α_{i0} is intercept of the regression point i , α_{ik} is the k th parameter of the regression point i , θ_i is the error, and Y is the dependent variable (water requirement of winter wheat).

The parameters in the GWR model change along with the spatial location and cannot be estimated by using the least squares method. Instead, the weighted least squares (WLS) method is used to estimate the parameters:

$$\alpha_i^* = (X^T \omega_i X)^{-1} X^T \omega_i Y, \quad (3)$$

where, α_i^* is the matrix of the regression coefficients; X and Y are the matrices of the independent (all the affecting factors) and dependent (crop water requirement) variables, X^T is the transpose of X ; and ω_i is the weighted matrix of $n \times n$. Every element along the diagonal line is a function between the location of observed value (j) and the regression point (i). Each element is used to determine the effect of the observed values in different spatial locations ($j = 1, 2, \dots, n$) on the estimation of the parameter at the regression point i . If the off-diagonal element is 0, the matrix ω_i can be expressed as follows [28]:

$$\omega_i = \begin{bmatrix} \omega_{i1} & & & \\ & \omega_{i2} & 0 & \\ & & \ddots & \\ & 0 & & \ddots & \\ & & & & \omega_{in} \end{bmatrix}. \quad (4)$$

The parameter is estimated by the WLS method, and the weight is normally determined based on the geographical spatial location. Eq. (5) is chosen as the weight function:

$$\omega_i(j) = \begin{cases} \left[1 - \left(\frac{d_i(j)}{h} \right)^2 \right]^2, & d_i(j) < h, \\ \omega_i(j) = 0, & d_i(j) \geq h, \end{cases} \quad (5)$$

where, $\omega_i(j)$ represents the weight of the j th monitoring point at point i , $d_i(j)$ is the distance from the j th monitoring point to point i , and h is the bandwidth, which is the parameter used to control the model smoothening degree and determined by the cross-verification method because there exists a certain aggregation in the monitoring station in the study area.

2 Result and discussion

2.1 Global regression analysis

In this study, 15 affecting factors in three categories (weather; macro- and micro-terrain factors) are obtained from 67 stations in North China. ArcGIS is used to perform global regression analysis (GRA) between the 15 factors and the water requirement of winter wheat. The results are shown in Table 1.

Table 1 shows that the variance inflation factor (VIF) of most variables is >10 [20], although the adjusted R^2 is high, indicating that variable redundancy is found in this regression equation and there is collinearity among some variables. Therefore, the hypothesis that the variables are relatively independent is not valid.

To eliminate the correlation among the affecting factors, we performed PCA and GRA of the 15 factors as shown in Table 2.

As can be seen from Table 2, all of the VIFs of the variables are <10 , which is revealed by PCA of the original variables. Collinearity among those variables is eliminated by using PCA. Akaike information criterion (AIC) decreases from 639.40 to 616.48, whereas the adjusted R^2 increases from 0.5776 to 0.6999, indicating that the effect of regression is improved after the factors are transformed.

To determine if there exists autocorrelation of the residual errors between these two regression models, Moran's I is used to calculate the autocorrelation coefficient. The value of Moran's I ranges from -1 to $+1$, in which $I > 0$ indicates a positive correlation, whereas $I < 0$ indicates a negative correlation. Thus, the greater the correlation in spatial distribution, the greater Moran's I absolute value results and the lower correlation in the spatial distribution. Namely, there exists spatially aggregated distribution. The spatial distribution becomes random when the Moran's I value approaches 0. In addition, a "significance test" on I value is

Table 1 Estimated parameter values of global regression analysis ($n = 67$)^{a)}

Variable	Regression coefficient	Standard error	<i>T</i>	<i>P</i>	VIF
Constant term	194.16	516.16	0.38	0.71	–
Longitude	0.23	3.08	0.07	0.94	9.18
Latitude	2.27	6.30	0.36	0.72	14.03
Elevation	–0.01	0.04	–0.38	0.71	21.24
Gradient	0.69	2.96	0.23	0.82	2.44
Aspect	–2.15	5.08	–0.42	0.67	1.45
Shading degree	123.38	120.09	1.03	0.31	1.32
Average temperature during growing period	21.98	38.46	0.57	0.57	423.27
Wind speed	23.89	9.32	2.56	0.013356*	9.32
Sunshine hours	0.07	0.05	1.38	0.17	15.53
Minimum temperature	–18.75	39.05	–0.48	0.63	682.14
Maximum temperature	6.34	10.63	0.60	0.55	37.15
Relative humidity	–4.31	2.02	–2.13	0.037695*	12.90
Precipitation	0.01	0.16	0.04	0.97	6.88
Evaporation	–0.06	0.05	–1.32	0.19	8.43
Daily temperature range	–0.94	22.21	–0.04	0.97	178.35
Diagnostic result of regression analysis					
Number of samples	67	Number of variables		16	
Degree of freedom	51	AIC		639.40	
R^2	0.6736	Adjusted R^2		0.5776	

a) * represents the significant difference at 5% significance level; *T* is the *t* statistic of the regression; *P* is the *P* statistic of the regression; VIF is the variance inflation factor; AIC is the Akaike information criterion.

Table 2 Results of global regression analysis based on the principal component analysis ($n = 67$)^{a)}

Variable	Regression coefficient	Standard error	<i>T</i>	<i>P</i>	VIF
Constant terms	434.68	77.88	5.58	0.000001**	–
PCA1	–10.52	2.15	–4.88	0.000011**	3.28
PCA2	10.12	2.12	4.76	0.000017**	2.54
PCA3	–9.85	3.32	–2.96	0.004605**	2.79
PCA4	–5.52	3.15	–1.76	0.09	1.28
PCA5	4.55	6.29	0.72	0.47	3.80
PCA6	0.35	4.75	0.07	0.94	2.33
PCA7	–10.21	10.26	–1.00	0.32	2.98
PCA8	0.00	12.04	0.00	1.00	3.16
PCA9	4.64	12.84	0.36	0.72	3.06
PCA10	–21.86	12.70	–1.72	0.09	1.66
PCA11	–12.55	13.10	–0.96	0.34	1.56
PCA12	10.07	14.60	0.69	0.49	1.56
PCA13	47.66	17.55	2.72	0.009004**	1.48
PCA14	80.85	25.18	3.21	0.002289**	1.43
PCA15	111.86	65.07	1.72	0.09	1.40
Diagnostic result of regression analysis					
Number of samples	67	Number of variables		16	
Freedom	51	AIC		616.48	
R^2	0.7682	Adjusted R^2		0.6999	

a) PCA1 to PCA15 correspond to the 1st to 15th principal component of the 15 variables in Table 1 through PCA; ** represents the significant difference at 1% significance level.

conducted and there is a significant correlation in certain distributions in the study area at a significance level of 5% (or 1%), i.e., there is spatial autocorrelation between spatial elements in the research scope if $Z(I)$ is >1.96 (or 2.58). If $Z(I)$ is between -1.96 (or -2.58) and 1.96 (or 2.58), then the correlation of a certain distribution in the research scope is insignificant and the spatial autocorrelation is weak. Moreover, if $Z(I)$ is <-1.96 (or -2.58), then there exists a negative spatial autocorrelation in that certain distribution in the research scope. Table 3 shows the autocorrelation of the residual error in these two regression analyses.

As can be seen in Table 3, Moran's I value is 0.23 and $Z(I)$ value is $2.67 > 2.58$ prior to data processing, indicating that there exists a significant autocorrelation in the residual errors. After PCA, Moran's I and $Z(I)$ values decrease to 0.14 and 1.64 , respectively, indicating that the autocorrelation of the residual error is reduced and the interpretability of the regression equation is improved.

Table 2 shows the effect among the five factors (PCA1, PCA2, PCA3, PCA13, and PCA14) is significant at a significance level of 1%. To reduce the calculation workload, five principal factors are chosen as the factors that affect crop water requirement. Regression analysis is then conducted through GRA by considering water requirement as the dependent variable to compare with the regression coefficients of all of the variables as show in Table 4.

As can be seen in Table 4 the interpretation of the regression model does not significantly change, although 10 independent variables are removed. Neither AIC nor adjusted R^2 changes significantly. Moran's I and $Z(I)$ values of the residual errors increase, indicating a positive correlation

between residual errors. Therefore, further analysis through the GWR model is needed.

2.2 GWR analysis

GWR analysis of the five principal components as the independent variables is conducted by using the GWR tool in the ArcGIS statistical analysis software toolkit. "ADAPTIVEK" is chosen as the kernel function given the aggregation of the monitoring points. The optimized number of point neighborhood is 38 after several tests and bandwidth verification are conducted. To examine the stability of the regression coefficients in space, spatial autocorrelation analysis of constant terms and regression coefficients of the five factors is conducted (Table 5).

Table 5 shows that Moran's I values of the GWR coefficients are >0 and the $Z(I)$ values are >2.58 , indicating the strong spatial autocorrelation between the coefficients at a 1% significance level. Namely, spatial variability exists among these regression coefficients. Therefore, the quantitative effect of the different factors on crop water requirement in different spatial locations can be taken into account when GWR is used. Only the changing trend in the constant terms and PCA1 regression coefficients in space are analyzed due to space limitation (Figure 1(a) and (b)).

The constant terms reflect the basic water requirement distribution of winter wheat, which is mainly determined by energy factors. And temperature is the main energy factor closely related to the solar radiation that affects the crop metabolism. In particular, a higher metabolic rate within a certain temperature range results in higher soil evaporation rate and water requirement. The spatial distribution trend of the constant terms in Figure 1(a) is consistent with that of the average temperature during the growing season of winter wheat in Figure 2(a), i.e., spatial distribution is higher in the southeast and lower in the northwest, whereas higher in the plains and lower on the mountains. This changing trend of the constant terms is not always consistent with that of

Table 3 Autocorrelation analysis of residual error in the regression analysis before and after PCA of 15 variables ($n = 67$)

Item	Moran's I	$Z(I)$	P
Before PCA of 15 variables	0.23	2.67	0.0076
After PCA of 15 variables	0.14	1.64	0.1000

Table 4 Comparison of the estimated value in the global regression model ($n = 67$)

Item	AIC	R^2	Adjusted R^2	Residual error Moran's I	Residual error $Z(I)$
15 variables (regression analysis)	616.48	0.7682	0.6999	0.14	1.64
5 variables (regression analysis)	616.99	0.6951	0.6593	0.16	1.89

Table 5 Analysis of spatial variability of the regression coefficients of constant terms and five principal component factors ($n = 67$)

Item	Constant	PCA1 coefficient	PCA2 coefficient	PCA3 coefficient	PCA13 coefficient	PCA14 coefficient
Moran's I	0.891221	1.062763	0.764856	0.734611	0.887527	0.777675
Expected index	-0.015152	-0.015152	-0.015152	-0.015152	-0.015152	-0.015152
Variance	0.008767	0.008863	0.008837	0.008778	0.008868	0.008792
$Z(I)$	9.680058	11.449928	8.297617	8.002719	9.585455	8.455554
P	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

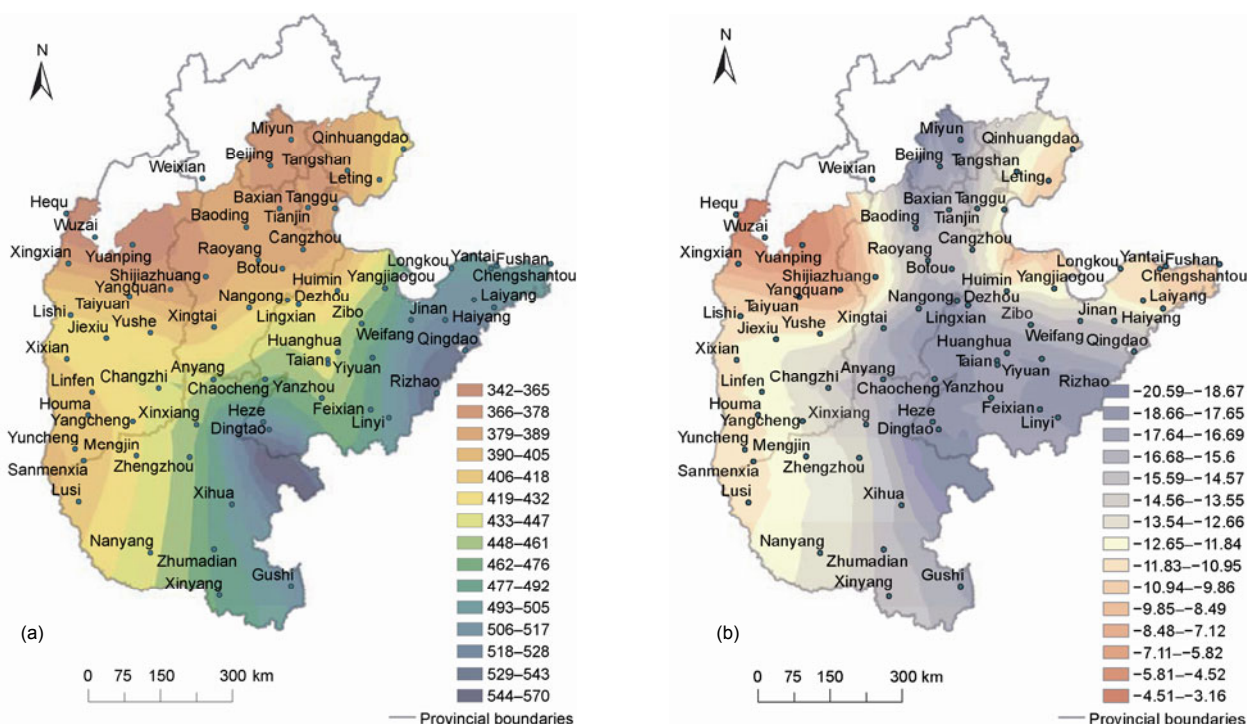


Figure 1 Spatial distribution of main parameter in the geographically weighted regression model. (a) Constant terms; (b) PCA1 regression coefficient.

the average temperature.

Figure 1(b) shows the regression coefficient of PCA1 is negative, indicating that the crop water requirement decreases with the increase in PCA1. A previous study reveals that PCA1 corresponds to the comprehensive factor of moisture and heat factors [19], i.e., rainfall and relative humidity contribute positively, whereas sunshine duration and evaporation contribute negatively to PCA1. Crop water requirement is low if the rainfall and relative humidity are high in a given place. The absolute value of the regression coefficient is high in the plains on the northern and the southern parts, followed by that in the northeast, and the lowest value in the northwest, i.e., the water requirement of winter wheat in the southern part decreases greatly compared with that in the northwestern part as PCA1 increases. This result is mainly attributed to the occurrence of higher rainfall in the southern part than that in the northwestern part of North China during the growing season of winter wheat. To verify this result, the spatial distribution of the sunshine duration during the growing season of winter wheat is made into Figure 2(b). As can be seen in Figure 2(b) the sunshine duration is shorter in the southern part compared with that in the northwestern part of North China, indicative of more rainfall occurs or cloudy and humid conditions at the same humidity factor in the southern part compared with those in the northeastern part. Therefore, PCA1 elicits a significant effect on crop water requirements.

The results of GRA and GWR are summarized in Table 6 to illustrate the effect of the weighted regression model.

As can be seen in Table 6 the AIC value decreases significantly from 616.99 to 274.64, whereas the adjusted R^2 increased from 0.6593 to 0.7616, Moran's I and $Z(I)$ values decreased from 0.16 to 0.09 and 1.89 to 1.14, respectively. The autocorrelation of the residual error was also decrease and the interpretability of the model enhances.

2.3 Spatial distribution of the water requirement of winter wheat in North China

Based on the spatial distribution of the factors that affect the water requirement and their regression coefficients, the ArcGIS software for grid calculation is used to obtain the spatial distribution of crop water requirements in the study area (Figure 3(a)).

Figure 3(b) illustrates the trends of the regional variables to prove the rationality of Kriging method. The general trends (Figure 3(b)) of the water requirement of winter wheat in North China are consistent with those in Figure 3(a), i.e., the water requirement is high in the northern part and low in the southern part. In the basin of Changzhi, Shanxi Province, a high water requirement is observed. The water requirement of winter wheat is also high in Beijing, Baoding, Botou, Lingxian County, and Taian, whereas the water requirement is low in Shijiazhuang, Xingtai, and Chaocheng. This phenomenon is partly due to the meteorological and terrain factors. For example, compared with the neighboring places, more rainy days and shorter sunshine duration are experienced in Shijiazhuang and Xingtai, among others, on the plains at the base of Taihang Mountain,

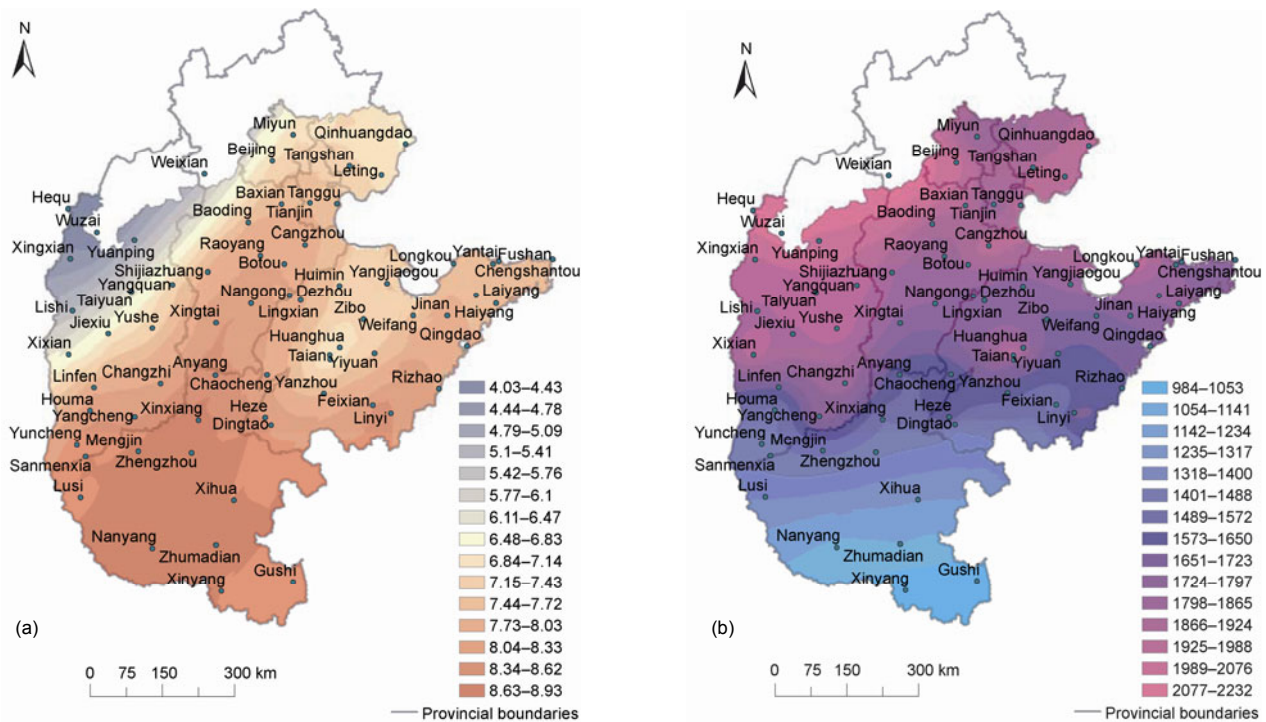


Figure 2 Spatial distribution of main meteorological elements during the growing season of winter wheat (h). (a) Average temperature(°C); (b) sunshine duration (h).

Table 6 Comparison of the global regression analysis and geographically weighted regression model ($n = 67$)

Item	AIC	R^2	Adjusted R^2	Residual error Moran's I	Residual error $Z(I)$
15 variables (GRA)	616.99	0.6951	0.6593	0.16	1.89
5 variables (GWR)	274.64	0.7922	0.7616	0.09	1.14

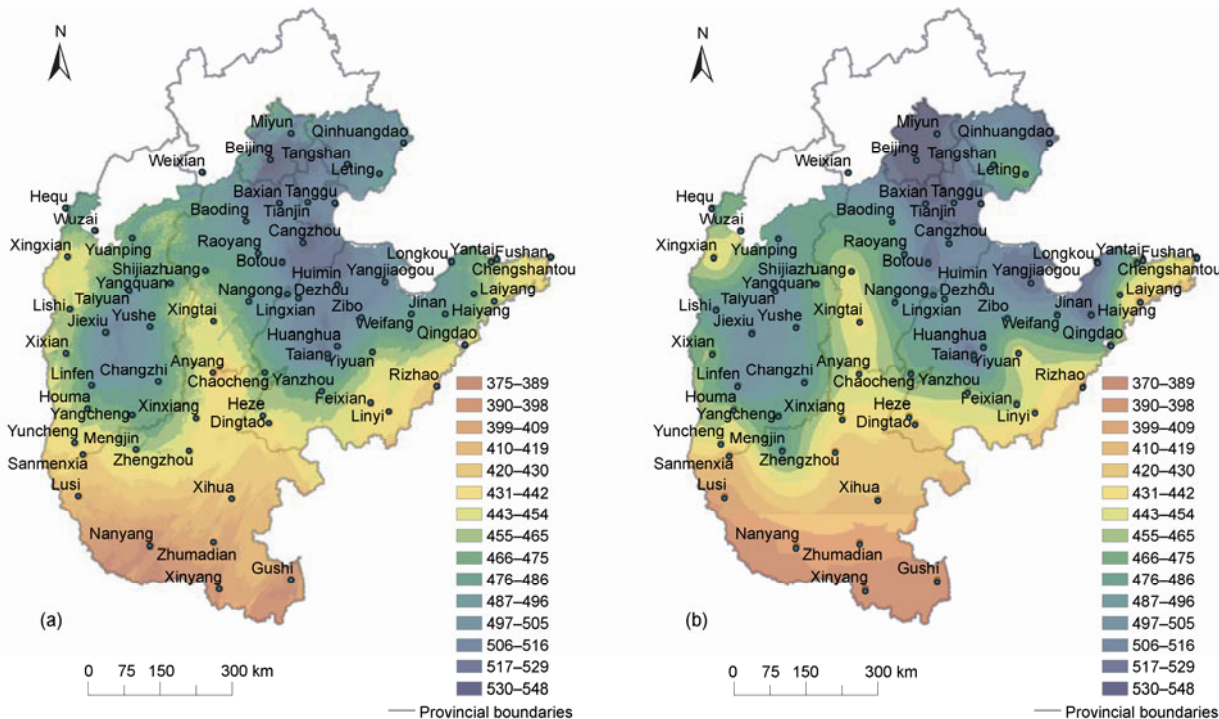


Figure 3 Spatial distribution of the average water requirement of winter wheat. (a) PCA+GWR (mm); (b) Kriging (mm).

where low water requirement of winter wheat is observed because Taihang Mountain provides shade in the area.

Generally, mean square error of prediction which measures the average square difference between the true crop water requirement and its estimate in the validation points (MSE), mean bias error (MBE), mean absolute error (MAE), scaled mean square error of prediction which measures the average square difference between the observed crop water requirement and its estimate crop water requirement divided by the observed crop water requirement in the Nv-validation points (s-MSE), and linear correlation coefficient (CC) are used to assess interpolation methods [29]. The comparison of the spatial processing of these two methods is shown in Table 7.

Table 7 shows that 4 out of 5 indexes in the method (PCA+GWR) proposed in this study are better than those in Kriging, whereas the mean variation (s-MSE) is less than that in Kriging. Thus, the proposed method is of significant practical value in estimating the regional water requirement of crops. Furthermore, Kriging method exhibits some difficulties in expressing the spatial variation in relatively small areas (such as near Beijing) because it does not consider the effect of the geographic coordinates (macro-terrain factor) or the spatial change in micro-terrain and meteorological factors. The Kriging method also exhibits an inherent smoothening effect [30]. Taking into account the macro- and micro-terrain as well as meteorological factors, this method (PCA+GWR) can accurately express the spatial change in the crop water requirement caused by changes in the micro-terrain and the meteorological factors. This method can also provide detailed information about the water requirement of winter wheat, when the micro-terrain factor or the meteorological factor suddenly changes.

3 Conclusion

This study combines PCA and GWR to eliminate the collinearity of the factors that affect the crop water requirement, reduce the autocorrelation of the residual errors in global regression equation, and improve the interpretability of regression analysis. PCA is initially used to eliminate the collinearity of the affecting factors. GWR is then conducted by analyzing the variability of the spatial data to estimate and illustrate the spatial distribution of the water requirement of the winter wheat in North China. Results show that PCA of the affecting factors can effectively eliminate the collinearity of the affecting factors, increase the interpretability of the global regression equation, and improve the appraisal

indexes of the regression equation. However, spatial autocorrelation of the residual errors still exists. AIC and the autocorrelation of the residual errors significantly decrease, whereas R^2 significantly increase compared with GRA when GWR is used. Therefore, GWR can be used to elucidate the spatial difference of the effect of the factors on crop water requirement when these factors are used. The spatial distribution of the water requirement of the winter wheat in North China and the possible causes are analyzed based on the changes in the regression coefficient and the spatial changes in these factors. In contrast to the traditional global regression, the method proposed in this study can eliminate the correlation of the affecting factors, improve data processing efficiency, reduce the spatial autocorrelation of residual errors, increase the interpretability of the regression model. It can also reflect the changes in crop water requirement with the spatial location. In addition, compared with Kriging, in which only the effect of the space coordinates on crop water requirement is considered, the proposed method can clearly show the effect of the factors in different spatial locations on crop water requirement, and thus provide reference for the estimation of the regional crop water requirement.

This work was supported by the National Basic Research Program of China (2006CB403406), the National Natural Science Foundation of China (51079154) and the National High-Tech Research & Development Program of China (2011AA100502).

Table 7 Comparison of spatial processing

Method	MSE	MBE	MAE	s-MSE	CC
PCA+GWR	16.31	0.683	12.66	0.0012	0.897
Kriging	21.88	0.286	21.88	0.0021	0.835

- Chen Y M, Guo G S, Wang G X, et al. Water Requirement and Irrigation of Main Crops in China (in Chinese). Beijing: Water Power Press, 1995. 49
- Sun J S, Xiong Y Z, Kang S Z. Progress and research method in evapotranspiration in cropland (in Chinese). J Irrig Drain, 1993, 12: 36–38
- Xu D, Liu Y. A survey of measurement and estimation of crop evapotranspiration (in Chinese). J Irrig Drain, 1997, 16: 54–59
- Rana G, Katerji N. Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: A review. Eur J Agron, 2000, 13: 125–153
- Wang X Y. Study of the estimating methods for evapotranspiration in farmland (in Chinese). Sys Sci Compr Stud Agr, 2003, 19: 81–84
- Wang J L, Sun J S, Fu M J, et al. Existing problems and solving methods of regional crop water requirements evaluation (in Chinese). Water Sav Irrig, 2005, 30: 4–7
- Amegee K Y. Application of geostatistics to regional evapotranspiration. Dissertation for Doctoral Degree. Corvallis: Oregon State University, 1985. 176
- Yuan X, Li E Y. Spatial variability of reference crop evapotranspiration (in Chinese). J Hydraul Eng, 1990, 21: 33–37
- Harcum J B, Loftis J C. Spatial interpolation of Penman evapotranspiration. T ASABE, 1987, 30: 129–136
- Gu S X, He D M, Cui Y L, et al. Spatial variability of irrigation factors and their relationships with “corridor-barrier” functions in the Longitudinal Range-Gorge Region. Chin Sci Bull, 2007, 52(Suppl II): 33–41
- Hashmi M A, Garcia L A. Spatial and temporal errors in estimating regional evapotranspiration. J Irrig Drain E, 1998, 124: 108–114
- Daleziosa N R, Loukasb A, Bampzelisc D. Spatial variability of reference evapotranspiration in Greece. Phys Chem Earth, 2002, 27: 1031–1038

- 13 Yue T X, Liu J Y. A digital model for multi-sources information fusion (in Chinese). *World Sci-Tech R&D*, 2001, 23: 1–4
- 14 Tong L, Kang S Z, Zhang L. Temporal and spatial variations of evapotranspiration for spring wheat in the Shiyang river basin in northwest China. *Agr Water Manage*, 2007, 87: 241–250
- 15 Ma R H, Huang X Y, Zhu C G. Knowledge discovery with ESDA from GIS database (in Chinese). *J Remote Sens*, 2002, 6: 102–107
- 16 Wu J S, Kang S Z, Wang J L, et al. Zoning of water-saving irrigation in China using principal components analysis and fuzzy-C-means (in Chinese). *Trans Chin Soc Agric Eng*, 2004, 20: 64–68
- 17 Mathes S E, Rasmussen T C. Combining multivariate statistical analysis with geographic information systems mapping: A tool for delineating groundwater contamination. *Hydrogeol J*, 2006, 14: 1493–1507
- 18 Zhou L Z, Ma Y, Li D Q. Spatial distribution patterns of Chinese gerbils (gerbilline) in relation to environmental factors (in Chinese). *Acta Zool Sin*, 2001, 47: 616–624
- 19 Wang J L, Sun J S, Song N, et al. Analysis of dominant factors of winter wheat water requirements based on GIS and PCA (in Chinese). *Eng J Wuhan Univer*, 2009, 42: 640–643
- 20 Fotheringham A S, Charlton M E, Brunsdon C. The geography of parameter space: An investigation into spatial non-stationarity. *Int J Geogr Inf Syst*, 1996, 10: 605–627
- 21 Su F L. Analysis on spatial factors for county economic development based on GWR model—take Liaoning Province counties as an example (in Chinese). *Acade Forum*, 2005, 28: 81–84
- 22 Xuan H Y, Li S P, Liu S Q. Analysis on the relationship between the precipitation on the one hand and longitude, latitude and altitude above sea level on the other (in Chinese). *J Gansu Sci*, 2006, 18: 26–28
- 23 Chang L F, Su M D. Using the geographically weighted regression to modify the residential flood damage function. In: *World Environmental and Water Resources Congress*, 2007. 1–25
- 24 Ghosh D, Manson S M. Robust principal component analysis and geographically weighted regression: Urbanization in the Twin cities Metropolitan area of Minnesota. *URISA Jnl*, 2008, 20: 15–26
- 25 Zhang L F, Gu J T, Lin J N. Using geographically weighted regression to build the relationship with irrigation rate and effecting factors. *J Chin Agri Eng (Taiwan)*, 2006, 52: 73–82
- 26 Zhang L S, Wen H Q, Cheng T L, et al. Ecological regionalization of wheat cultivation in Shanxi Province of China (in Chinese). *Chin J Eco-Agri*, 2010, 18: 410–415
- 27 Zhao G C. Study on Chinese wheat planting regionalization (in Chinese). *J Triticeate Crops*, 2010, 30: 886–895
- 28 Tan W Z, Wang J M, Liu M L. Spatial nonstationarity of geographically weighted regression analysis of spatial data (in Chinese). *J Liaoning Normal Univer (Nat Sci Ed)*, 2005, 28: 476–479
- 29 Piazza A D, Conti F L, Noto L V, et al. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *Int J Appl Earth Obs Geoinf*, 2011, 13: 396–408
- 30 Shi W J, Liu J Y, Du Z P. High accuracy surface modeling of soil properties based on geographic information (in Chinese). *Acta Geogr Sin*, 2011, 66: 1574–1581

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.